

4. Skalenerträge

„Wo immer etwas falsch ist, ist es zu groß“

Leopold Khor

Mergers und Aquisitions werden vom Management meist durch Synergien und große damit verbundene Einsparungspotentiale motiviert. Viele Studien in der Industrieökonomie zeigen allerdings, dass Unternehmensverschmelzungen oft den Aktienkurs dämpfen und die versprochenen Einsparungen durch Reibungsverluste während der Integrationsphase verloren gehen (Moeller – Schlingemann – Stulz, 2004; Tichy, 2001). Kleine und überschaubare Unternehmen leiden nicht unter Informations- und Koordinationsproblemen und können daher besser auf Marktänderungen reagieren.

Größenvorteile in der Produktion von Gütern oder Dienstleistungen sind ein wichtiges Thema für die strategische Entscheidung des Managements über die Kapazität eines Betriebes und gleichzeitig ein altes Thema der Wettbewerbspolitik. Größenvorteile entstehen, wenn es einem Unternehmen gelingt, durch die Ausweitung der Produktionskapazität den Output überproportional zu steigern. Die wirtschaftliche Konsequenz so genannter steigender Skalenerträge sind mit dem Produktionsniveau fallende Durchschnittskosten. Aus der Sicht des Managements kann also durch die Wahl der Betriebsgröße der Durchschnittskostensatz gesteuert werden. Die Beschleunigung von Investitionsplänen begleitet von Preissenkungen und damit einer Ausweitung des Absatzes ist bei steigenden Skalenerträgen profitabel, weil niedrigere Durchschnittskosten anfallen.

Besonders interessant ist die Eigenschaft des Produktionsprozesses in einem neu entstehenden Markt. Durch die Wahl der Produktionskapazität kann der Ersteintretende in einem Markt potenzielle weitere Markteintritte abschrecken, und schafft sich damit ein Freiraum in der Festlegung des Absatzpreises. Ein Nebeneffekt von Größenvorteilen in der Produktion betrifft die Anzahl potenzieller Anbieter im betreffenden Markt. Wenn die Betriebsgröße, bei der die Durchschnittskosten ihr Minimum erreichen, fast gleich groß ist wie das gesamte Marktvolumen, kann nur ein Unternehmen diesen Markt profitabel versorgen. In diesem Fall spricht man von einem natürlichen Monopol (Varian, 1993).

Die durch New Economy Welle neu geschaffenen Märkte sind ein gutes Beispiel für Vorteile eines First Movers bei steigenden Skalenerträgen. Amazon, E-bay und Google konnten mit ihrem Geschäftsmodell durch die rasche internationale Expansion sowohl die kritische Anzahl der Nutzer überschreiten als auch die Geschäftsmöglichkeiten für nachfolgende Unternehmen einengen. Das Profitpotential aus steigenden Skalenerträgen auf elektronischen Märkten hatte in den späten 1990er Jahren dementsprechend hohe Bewertungen der Aktien solcher Unternehmen zur Folge.

Die vielen Pleiten am Ende der Internet-Bubble zeigen aber, dass nur wenige Märkte die notwendigen Voraussetzungen zum Ausnutzen steigender Skalenerträge haben. Aus der Sicht der Wettbewerbspolitik ist in einem monopolistisch oder oligopolistisch strukturierten Markt die gesamtwirtschaftliche Wohlfahrt kleiner als in einem Markt mit perfektem Wettbewerb. Daher gibt es in fast allen Ländern eine von zwei Formen des staatlichen Eingriffs. Entweder gibt es für einen Markt eine Regulierungsbehörde, die die dort aktiven Privatunternehmen beaufsichtigt und entsprechende Rahmenbedingungen festlegt (z. B. Telefon, Strom, Gas), oder die Produktion erfolgt direkt im öffentlichen Sektor durch Körperschaften oder ausgegliederte Unternehmen (z. B. öffentliche Verkehrsmittel, Wasserversorgung, Abfallentsorgung). Die Europäische Kommission versucht im Rahmen des Binnenmarktprojektes eine Harmonisierung der Produktionsbedingungen und einen stärkeren Wettbewerb auch in vormals verstaatlichten oder regulierten Industrien zu bringen.

Für die Wettbewerbspolitik sind zwei Informationen von zentraler Bedeutung. Erstens, gibt es in einem Markt steigende Skalenerträge oder nicht? Ein Regulator als ordnungspolitische Instanz oder die öffentliche Hand als direkter Produzent müssen also die Produktionsfunktion einer Industrie kennen. Zweitens, sind die Eigenschaften der Nachfragefunktion in diesem Markt relevant. Aus der Kombination beider Informationen können die Vorgaben für einen Wirtschaftsbereich festgelegt werden, z. B. die Durchleitungsgebühren, die Absatzpreise oder auch die Produktionsmengen.

Wir werden uns in diesem Abschnitt nur mit der Angebotsseite der Energiewirtschaft beschäftigen. Wodurch entstehen eigentlich steigende Skalenerträge in einer Industrie?

Für Größenvorteile gibt es rein technische Ursachen oder wirtschaftliche Gründe durch hohe Anfangsinvestitionen vor Beginn der eigentlichen Produktion. Zur praktischen Veranschaulichung technisch bedingter Skalenerträge wird oft die Auswirkung einer Erweiterung des Rohrdurchmessers auf eine Pipeline angeführt. Bei einer Verdoppelung des Durchmessers von 50cm auf 100cm steigt zwar das Material zur Erstellung der Pipeline um das Doppelte (Kreisumfang), das Durchgangsvolumen der Röhre nimmt hingegen um den Faktor vier zu (Kreisfläche). Daher steigt bei einer Vervielfachung des Rohrdurchmessers die Menge an gepumptem Öl oder Gas überproportional zum verwendeten Material für die Pipeline.

Das Beispiel einer Pipeline macht gleichzeitig deutlich, dass steigende Skalenerträge in der Regel ein lokales Phänomen sind. Wenn der Rohrdurchmesser tatsächlich immer wieder verdoppelt werden würde, müsste die Konstruktion letztendlich unter ihrem eigenen Gewicht zusammenbrechen. Zusätzlich müssen Pumpen, Zulauf- und Ablaufspeicher usw. auf die Durchlaufmenge ausgerichtet werden. Für die erwartete Transportleistung bedeutet das, dass im Fall einer Pipeline ein technisch optimales Produktionsniveau existiert, welches nicht überschritten werden kann. Weiters

kann technischer Fortschritt die Relation zwischen allen diesen Größen verschieben.

Produktionsfunktionen haben gewöhnlich sinkende Grenzerträge, d. h. wenn die eingesetzte Menge eines Produktionsfaktors steigt, wird die zusätzlich produzierte Menge immer geringer. Wie passen sinkende Grenzerträge mit steigenden Skalenerträgen zusammen?

In der mikroökonomischen Analyse unterscheidet man zwischen kurz- und langfristiger Analyse. Die kurze Frist zeichnet sich dadurch aus, dass ein oder mehrere Entscheidungsvariablen durch das Management nicht verändert werden können. Der Bestand an Gebäuden oder Maschinen in einem Unternehmen ist z. B. kurzfristig gegeben. Unter diesen Umständen ist eine ceteris paribus Analyse sinnvoll, d. h. wir konzentrieren uns auf die marginale Wirkung eines Eingriffs in das System und interessieren nur z. B. nur für den Grenzertrag eines Produktionsfaktors. Damit sind die Ergebnisse unserer Analyse von der Ausgangslage abhängig und nicht allgemein gültig.

Betriebswirtschaftliche und wirtschaftspolitische Maßnahmen bestehen aber oft aus der gleichzeitigen Änderung aller Rahmenbedingungen im selben Ausmaß, z. B. werden zur Ausweitung der Produktionskapazität eines Unternehmens sowohl die Gebäude als auch die Maschinen und Arbeitnehmer aufgestockt. Solche Schritte kann ein Unternehmen aber nicht kurz-, sondern nur mittel- bis langfristig setzen, weil Umbauten, Installationen, Personalsuche sowie Einschulungen Zeit kosten. Das Konzept der Skalenerträge wird zum Verständnis der Reaktion des Produktionsvolumens auf die Veränderung aller Inputs im gleichen Ausmaß verwendet.

Die Konzept der Skalenerträge kann mit einer einfachen Fragestellung dargestellt werden (Varian, 1993 Kap. 17): Um wie viel steigt die Produktion eines Unternehmens, wenn wir gleichmäßig alle Inputs um denselben Faktor erhöhen? Eine nahe liegende Antwort wäre, dass eine Verdoppelung der Inputs zur zweifachen Produktionsmenge führt. Wenn ein Zwillingbetrieb neben das Originalwerk gestellt wird, sollte die Produktion dieser völlig gleichartigen Betriebe genau doppelt so hoch sein. Diesen Fall bezeichnen wir als Produktion mit konstanten Skalenerträgen. Konstante Skalenerträge unterstellen, dass ein Unternehmen in der Lage sein sollte, das aktuelle Produktionsniveau mit Hilfe der um den Faktor k vervielfachten Inputs um denselben Faktor k zu vermehren.

Wenn wir eine Produktionsfunktion $f(\cdot)$ für den Output Y mit zwei Inputs x_1 und x_2 ganz allgemeinen als

$$Y = f(x_1, x_2),$$

definieren, können wir konstante Skalenerträge unter der Annahme verdoppelter Inputs mathematisch einfach als

$$2f(x_1, x_2) = f(2x_1, 2x_2),$$

darstellen. Dieser Ansatz kann ganz einfach verallgemeinert werden, indem wir die Verdoppelung durch eine Vervielfachung um einen beliebigen Faktor k ersetzen:

$$kf(x_1, x_2) = f(kx_1, kx_2).$$

Wir haben jetzt gesehen, wie eine Technologie mit konstanten Skalenerträgen aussieht, wie kann man eigentlich eine Produktionsfunktion mit steigenden Skalenerträgen darstellen?

Das ergibt sich eigentlich schon durch die Bezeichnung „steigend“. Darunter versteht man eine Produktionsfunktion, die auf eine gleichzeitige Steigerung aller Inputs um einen Faktor k mit einer überproportionalen Erhöhung der Produktion reagiert:

$$f(kx_1, kx_2) > kf(x_1, x_2) \quad \text{für } k > 1.$$

Wie schon gesagt, steigende Skalenerträge sind in diesem Fall rein durch die Technologie, also die Produktionsmöglichkeiten erklärt! Wenn wir wirtschaftliche Gesichtspunkte in die Analyse einbringen wollen, müssen wir Annahmen über die Verhaltensweise der Unternehmen treffen. Ein nahe liegendes Verhalten für ein Management ist Kostenminimierung.

Wir wissen aber aus dem Dualitätsprinzip, dass die aus einer Kostenminimierung abgeleitete Kostenfunktion im Grunde dieselbe Information enthält wie die Produktionsfunktion. Damit fasst die Kostenfunktion alle wirtschaftlich relevanten Eigenschaften einer Produktionsfunktion zusammen. Aus einer Kostenfunktion kann die dazu gehörende Technologie abgeleitet werden. Unter bestimmten Regularitätsbedingungen gibt es sogar einen eindeutigen Zusammenhang zwischen Kosten- und einer bestimmten dazu gehörenden Produktionsfunktion (Varian, 1992). Das Dualitätsprinzip ist für die empirische Analyse sehr wichtig, weil viele Probleme in der direkten Analyse der Produktionsfunktion nicht oder schwierig untersucht werden können, während Kostenfunktionen einen leichten Ansatz bietet.

Die Kostenfunktion für ein Unternehmen ergibt sich aus der Lösung des Optimierungsproblems mit einer Nebenbedingung:

$$\begin{aligned} \min_{x_1, x_2} p_1 x_1 + p_2 x_2, \\ \text{s.t. } f(x_1, x_2) = Y, \end{aligned}$$

wobei p_1 und p_2 die Preise der beiden Produktionsfaktoren sind. Die daraus folgende Kostenfunktion $c(p_1, p_2, Y)$ zeigt die minimalen Kosten zu denen ein Produktionsniveau von Y erreicht werden kann (Varian, 1993 Kap. 19). Die Auswirkung konstanter Skalenerträge auf die Kostenfunktion ist klar erkennbar, wenn wir von der Produktion einer Einheit von Y ausgehen. In diesem Fall sieht die Kostenfunktion folgendermaßen aus: $c(p_1, p_2, 1)$. Der billigste Weg zur Produktion von Y Einheiten erfordert unter konstanten Skalenerträgen einfach die Vervielfachung der Inputs um Y gegenüber dem Fall mit einer Output-

Einheit. Die Kostenfunktion ist dann $c(p_1, p_2, Y)$ und die Kosten steigen linear mit dem Produktionsniveau.

Wenn wir steigende Skalenerträge annehmen, steigen die Kosten unterproportional zur Produktionsmenge, d. h., wenn ein Unternehmen die doppelte Produktionsmenge erzielen möchte, steigen die Kosten um weniger als das Doppelte.

Voraussetzung dafür ist selbstverständlich, dass das Unternehmen keinen Einfluss auf die Inputpreise hat!

Besonders anschaulich wird die Wirkung von Skalenerträgen auf die Kostenfunktion, wenn wir die Durchschnittskosten (average costs) $AC(Y)$ betrachten. Dazu dividieren wir einfach die Kostenfunktion durch die Anzahl der produzierten Einheiten Y :

$$AC(Y) = \frac{c(p_1, p_2, Y)}{Y},$$

und nutzen unser Wissen, dass unter konstanten Skalenerträgen die Kostenfunktion folgende Form hat: $c(p_1, p_2, Y) = c(p_1, p_2, 1) Y$. Unter konstanten Skalenerträgen ist die Durchschnittskostenkurve also konstant und von der Anzahl der erzeugten Einheiten völlig unabhängig:

$$AC(Y) = \frac{c(p_1, p_2, 1)Y}{Y} = c(p_1, p_2, 1).$$

Steigende Skalenerträge sind mit sinkenden durchschnittlichen Produktionskosten bei einer Steigerung der Produktionsmenge verbunden. Es ist aber klar, dass die Durchschnittskosten in den seltensten Fällen ewig sinken. Üblicherweise gibt es ein Minimum der langfristigen Durchschnittskostenkurve, welches die langfristig optimale Produktionsmenge anzeigt.

Wir haben bisher nur allgemeine Produktions- und Kostenfunktionen $f(\cdot)$ bzw. $c(\cdot)$ benutzt. Eine sehr beliebte Form der Produktionsfunktion ist die Cobb-Douglas-Funktion, benannt nach ihren beiden Schöpfern Charles Cobb und Paul Douglas (1928). Dieser Funktionstyp hat sehr angenehme analytische Eigenschaften und ist in Anwendungen leicht umsetzbar. Marc Nerlove (1963) verwendet eine Cobb-Douglas Produktionsfunktion mit drei Inputs als Grundlage für seine Schätzung der Skalenerträge in der US Elektrizitätswirtschaft.

Marc Nerlove untersuchte die Eigenschaften der US-Elektrizitätsindustrie während seiner Militärzeit. Es war die erste Arbeit in der die dualen Eigenschaften der Produktions- bzw. Kostenfunktion empirisch umgesetzt wurden. Nerlove geht von einer allgemeinen Cobb-Douglas Produktionsfunktion für Elektrizitätsversorgungsunternehmen aus:

$$Y = a_0 x_l^{a_l} x_k^{a_k} x_f^{a_f} u,$$

wobei die Elektrizitätsproduktion Y in kwh gemessen wird. Strom wird mit Hilfe von Arbeit, x_l , Kapital, x_k , und Brennstoffen, x_f erzeugt. Der Lohnsatz, p_l , der Preis für Kapital, p_k , und die Brennstoffkosten, p_f , bestimmen die Kosten der Produktion. Die Skalierungsvariable a_0 entspricht einer Konstante und die Variable u beschreibt die Unterschiede in der Effizienz zwischen den Versorgungsunternehmen. Kostenminimierung unter der Nebenbedingung dieser Produktionsfunktion impliziert die folgende Kostenfunktion:

$$c = r \left(a_0 a_l^{a_l} a_k^{a_k} a_f^{a_f} \right)^{-1/r} Y^{1/r} p_l^{a_l/r} p_k^{a_k/r} p_f^{a_f/r} u^{1/r},$$

wobei

$$r = a_l + a_k + a_f,$$

das Ausmaß der Skalenerträge misst. Wenn die Summe der drei Elastizitäten gleich 1 ist, sind die Skalenerträge konstant. Bei $r > 1$ bestehen steigende Skalenerträge und für $r < 1$ weist die Kostenfunktion sinkende Skalenerträge auf. Durch Logarithmieren der Kostenfunktion kann man diesen komplizierten Ausdruck vereinfachen:

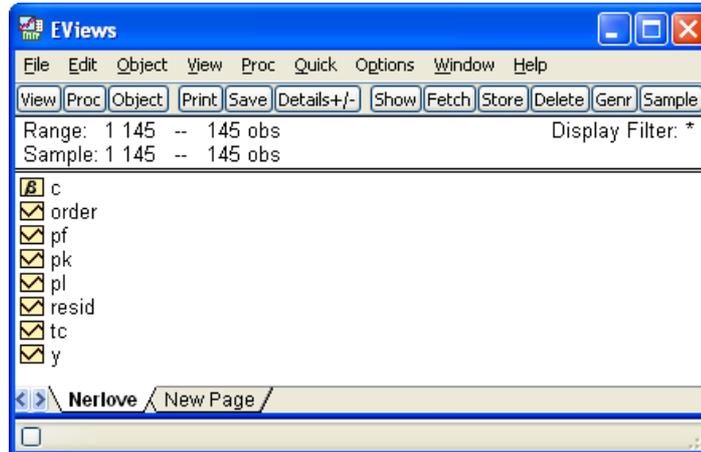
$$\log c = \log a_0 + \frac{1}{r} \log Y + \frac{a_l}{r} \log p_l + \frac{a_k}{r} \log p_k + \frac{a_f}{r} \log p_f + \frac{1}{r} \log u.$$

Eine Cobb-Douglas Produktionsfunktion erscheint mir für die Elektrizitätswirtschaft eigentlich ungeeignet. Sie unterstellt doch eine Substitutionsmöglichkeit zwischen den Inputs, die z. B. in einem kalorischen Kraftwerk unmöglich ist. Das Verhältnis zwischen Brennstoff, Kapital und Arbeit kann doch nicht beliebig verändert werden.

Deshalb verwendet Nerlove auch keine Betriebe als Beobachtungen, sondern ganze Versorgungsunternehmen, die über mehrere Betriebsstätten verfügen. In diesem Fall kann die Produktion schon zwischen den Standorten verschoben werden, sodass jene Technologie eingesetzt wird, die gerade die Kostenverhältnisse zwischen den Inputs am besten ausnützt.

4.1 Likelihoodschätzung, Konfidenzintervalle und die Delta Methode

Aus ökonomischer Sicht ist die obige Gleichung für eine erste Schätzung gut geeignet: sie ist linear und der Schätzer für die Skalenerträge ist direkt ablesbar als Kehrwert des Koeffizienten für $\log Y$. Die einzelnen Reihen stehen uns im file 04nerlove.wf1 zur Verfügung, wobei die Variable TC für die Kosten c steht, die anderen Variablen sind namensgleich. Die Variable ORDER bezeichnet hier die an sich unbedeutende Ordnungszahl aus einer größeren Gesamtheit an Betrieben. Nerlove hat vor allem nach Maßgabe der Datenverfügbarkeit 145 davon ausgewählt.



Eine erste Kleinst-Quadrateschätzung der obigen Gleichung ergibt folgendes Resultat:

Dependent Variable: LOG(TC)
 Method: Least Squares
 Sample: 1 145
 Included observations: 145

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.526503	1.774367	-1.987471	0.0488
LOG(Y)	0.720394	0.017466	41.24448	0.0000
LOG(PL)	0.436341	0.291048	1.499209	0.1361
LOG(PK)	-0.219888	0.339429	-0.647819	0.5182
LOG(PF)	0.426517	0.100369	4.249483	0.0000
R-squared	0.925955	Mean dependent var		1.724663
Adjusted R-squared	0.923840	S.D. dependent var		1.421723
S.E. of regression	0.392356	Akaike info criterion		1.000578
Sum squared resid	21.55201	Schwarz criterion		1.103224
Log likelihood	-67.54189	F-statistic		437.6863
Durbin-Watson stat	1.013062	Prob(F-statistic)		0.000000

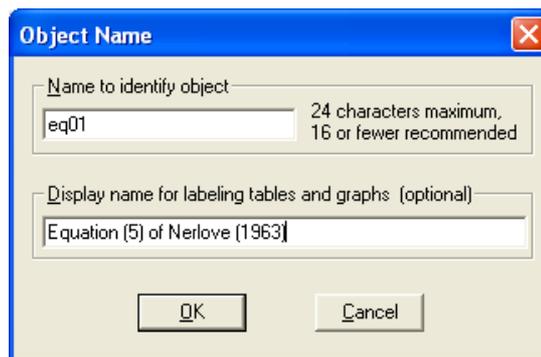
Wir erhalten also für den Koeffizientenschätzer von $\log Y$ einen signifikant von 0 verschiedenen Wert von etwa 0.72. De facto sind wir aber an ganz etwas anderem interessiert, nämlich zunächst an einem Schätzer für den Kehrwert dieses Koeffizienten, also \hat{r} und dann an einem Test, ob dieser Schätzer signifikant von 1 verschieden ist, was konstanten Skalenerträgen entspräche.

Um zu verstehen, warum die Informationen aus obiger Ergebnistabelle doch recht direkt dafür verwendbar wären, muss etwas weiter ausgeholt werden. Wir haben ja schon zur Ableitung des t-Tests die Annahme normalverteilter Störterme eingeführt. Diese Annahme lässt sich nun

für die Definition einer neuen Schätzmethode nützen, des so genannten Maximum-Likelihoodprinzips. Auf Basis der Normalverteilung lässt sich nun nämlich die gemeinsame Dichtefunktion aller Beobachtungen in einem linearen Regressionsmodell $y=X\beta+\varepsilon$ schreiben als $(2\pi\sigma^2)^{-n/2}\exp[-2\sigma^2(y-X\beta)'(y-X\beta)]$, was als Funktion in den unbekanntem Parametern bedingt auf die Daten $L(\beta,\sigma^2|y,X)$ interpretiert werden kann. Maximiert man nun diese Likelihood genannte Funktion (oder einfacher noch ihren Logarithmus) nach den Parametern, so erhält man dafür quasi jene Werte auf die die Daten am besten passen, die so genannten Maximum-Likelihood-(ML)-Schätzer. Der unter der Normalverteilungsannahme zu Stande gekommene ML-Schätzer für die Koeffizienten β entspricht übrigens exakt dem Kleinst-Quadrat-Schätzer (für eine Ableitung siehe etwa Hackl, 2005, S.51).

Mit Hilfe des ML-Prinzips lässt sich nun aber gut begründen, warum wir den Kehrwert $1/0.720394=1.388$ als Schätzer für r heranziehen können. Die inverse Transformation ist nämlich monoton im Parameterraum und hat deshalb keine Auswirkungen auf die Maximierung der ansonsten unberührten Likelihood-funktion. Das Gleiche gilt auch (annähernd) für den korrespondierenden statistischen Test, d.h. um zu überprüfen ob $\hat{r} = 1.388$ signifikant von 0 verschieden ist kann einfach der oben angegebene p-Wert herangezogen werden. Auch ein Test gegen den Wert 1 lässt sich einfach erstellen, zu diesem Zweck muss einfach eine neue t-Statistik aus $(\hat{\beta}_r-1)/\hat{\sigma}_\varepsilon\sqrt{(X'X)^{-1}_{rr}}$ erstellt werden.

In EViews lässt sich das über die Kommandozeile bewerkstelligen. Zunächst muss die obige Gleichung über Name benannt werden (z.B. eq01). Danach kann man über die Funktionen eq01.@coef(2) und eq01.@stderr(2) auf den Koeffizienten von log Y, respektive dessen Standardfehler zugreifen und sich dann direkt über die Verteilungsfunktion der t-Verteilung in der Kommandozeile über



`show 2*(1-@ctdist(@abs(eq01.@coef(2)-1)/eq01.@stderr(2),140))`

der p-Wert bestimmen lassen, welcher in vorliegendem Fall 0 ergibt, also zu einer Ablehnung der Hypothese konstanter Skalenerträge führte. Der Faktor 2 beziehungsweise die Verwendung der Absolutbetragsfunktion ergeben sich aus der Notwendigkeit einen zweiseitigen Test durchzuführen.

Gibt es nicht noch eine Möglichkeit Tests gegen andere Werte als 0 durchzuführen, ohne ständig die Testgrößen oder p-Werte neu bestimmen zu müssen?

Ja, man kann das so genannte Konfidenzintervall dafür verwenden. Dies bestimmt sich aus $[\hat{\beta}_i \pm \hat{\sigma}_\varepsilon\sqrt{(X'X)^{-1}_{ii}} t_{1-\alpha/2}]$, wobei $t_{1-\alpha/2}$ das t-Verteilungsquantil für einen α -Fehler mit entsprechenden Freiheitsgraden, nämlich $T-m-1$ bezeichnet.

In Eviews kann dies etwa für LOG(Y) wieder durch Eingabe in der Kommandozeile von

```
show eq01.@coef(2)-eq01.@stderr(2)*@qtdist(0.975,140)
show eq01.@coef(2)+eq01.@stderr(2)*@qtdist(0.975,140)
```

unter Verwendung der Verteilungsquantilfunktion @q?dist(1- α /2,Freiheitsgrade) errechnet werden, was ein Konfidenzintervall von [0.686, 0.755] ergibt.

Das Konfidenzintervall gibt also die Menge jener Werte an gegen die ein entsprechender t-Test mit einem α -Fehler von 5% zu keiner Ablehnung der Nullhypothese käme. Da in unserem Fall sowohl 0 als auch 1 außerhalb dieses Intervalls liegen, kommt es für beide Fälle zu einer Ablehnung.

Haben wir uns die Sache nicht doch ein wenig zu leicht gemacht, indem wir immer den Koeffizienten von LOG(Y) getestet haben – in Wirklichkeit sind wir doch an seinem Kehrwert also einer Transformation interessiert?

Das ist grundsätzlich richtig und muss beim Errechnen des Konfidenzintervalls natürlich auch berücksichtigt werden (im vorliegenden Fall bleibt dies jedoch – fast – ohne Konsequenzen). Zu diesem Zweck muss der Standardfehler (bzw. die Stichprobenvarianz) für den transformierten Schätzwert $\hat{r} = f(\hat{\beta}_1) = 1/\hat{\beta}_1$ bestimmt werden. Im vorliegenden Fall ergibt sich $\text{Vâr}(\hat{r}) = (\partial f / \partial \hat{\beta}_1)^2 \text{Vâr}(\hat{\beta}_1) = \text{Vâr}(\hat{\beta}_1) / \hat{\beta}_1^4$, was ein Konfidenzintervall von [1.322, 1.454] für r zur Folge hat.

Die obige Varianzformel ergibt sich aus der so genannten Delta-Methode zur Bestimmung der Standardabweichung von transformierten Zufallsgrößen. Diese motiviert sich aus einer Taylorreihenentwicklung um den geschätzten Erwartungswert der Größe, also $f(\beta) \approx f(\hat{\beta}) + (\beta - \hat{\beta}) \partial f / \partial \beta$ und damit auch für eine Varianz-Kovarianzmatrix $\text{Kov}(f, f') \approx \Delta \text{Kov}(\beta, \beta') \Delta'$, wobei Δ die Matrix mit den Gradientenvektoren $g_j = \partial f_j / \partial \beta$ für verschiedene Transformationen $f=(f_1, \dots, f_j, \dots)$ bezeichnet. Nach Einsetzen der Schätzwerte kann man dann in der Diagonale von $\text{Kôv}(f, f')$ die (Delta-)geschätzten Varianzen ablesen. Die Delta-Methode funktioniert umso besser, umso stärker die wahre Verteilung der Originalparameter um den Schätzwert konzentriert ist und liefert asymptotisch normalverteilte Zufallsgrößen. Im obigen, noch dazu einfachen skalaren Fall errechnet sich das korrigierte Konfidenzintervall mit Standardfehler $\sqrt{\text{Vâr}(\hat{r})} = 0.024$ über Eviews also aus

```
show 1/eq01.@coef(2)-eq01.@stderr(2)*@qnorm(0.975)/eq01.@coef(2)^2
show 1/eq01.@coef(2)+eq01.@stderr(2)*@qnorm(0.975)/eq01.@coef(2)^2
```

Im nicht-skalaren Fall ist diese Vorgangsweise in Eviews mühsam und mit Programmieraufwand verbunden. Papke und Wooldridge (2005) allerdings empfehlen einen einfachen Trick zur Bestimmung des Standardfehlers bei – auch nichtlinearen – Transformationen der Parameter. Man schätzt zuerst die Parameter untransformiert und evaluiert dann den Gradienten an dieser Stelle, anschließend transformiert man die Regressoren so, dass $\tilde{x}_i = [x_i - (\hat{\beta}_i / \hat{\beta}_k) x_k]$ für $i \neq k$ und $\tilde{x}_k = x_k / \hat{\beta}_k$ und regressiert den Regressanden darauf. Der gesuchte Standardfehler lässt sich dann aus der Tabelle in der Zeile für \tilde{x}_k ablesen. Da im vorliegenden Beispiel $\hat{\beta}=(0, 1/0.72, 0, 0, 0)$ gilt, braucht nur die Variable $\log Y$ entsprechend multipliziert zu werden:

Dependent Variable: LOG(TC)
 Method: Least Squares
 Sample: 1 145
 Included observations: 145

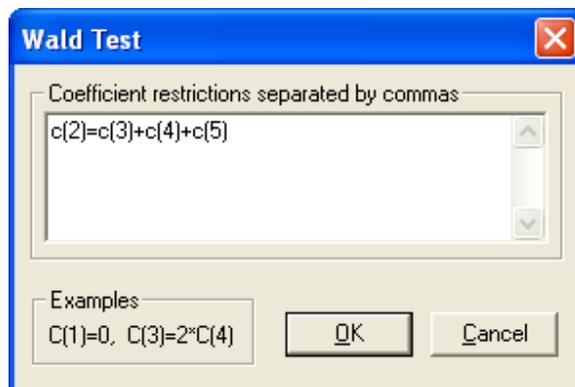
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.526503	1.774367	-1.987471	0.0488
LOG(Y)*0.720394	1.000000	0.024246	41.24448	0.0000
LOG(PK)	-0.219888	0.339429	-0.647819	0.5182
LOG(PL)	0.436341	0.291048	1.499209	0.1361
LOG(PF)	0.426517	0.100369	4.249483	0.0000
R-squared	0.925955	Mean dependent var		1.724663
Adjusted R-squared	0.923840	S.D. dependent var		1.421723
S.E. of regression	0.392356	Akaike info criterion		1.000578
Sum squared resid	21.55201	Schwarz criterion		1.103224
Log likelihood	-67.54189	F-statistic		437.6863
Durbin-Watson stat	1.013062	Prob(F-statistic)		0.000000

4.2 Restriktionen, der Wald-Test

Was bei der bisherigen Analyse auffällt ist, dass nicht berücksichtigt wurde, ob die logarithmierte Kostenfunktion die Homogenitätsbedingung an die Preise erfüllt, nämlich dass die Koeffizienten aller Preisvariablen sich auf 1 summieren müssen, damit wie aus dem Modell gefordert gilt, dass $r = a_l + a_k + a_f$.

Es handelt sich dabei um eine lineare Restriktion auf den Parametern, d.h. man kann sie allgemein in der Form $H\beta=h$ darstellen, wobei H eine $g \times m$ Matrix mit den Restriktionskoeffizienten für g Nebenbedingungen darstellt. Solche Nebenbedingungen können statistisch auf verschiedene Arten überprüft werden. Im Wesentlichen haben sich dafür drei Arten von Tests etabliert: der Wald-Test, der Lagrange-Multiplikator (LM) Test und der Likelihood-Ratio (LR) Test. Die Teststatistik für den Wald-Test basiert auf der Abweichung von der Restriktion an der Stelle des unrestringierten Schätzers, also $d = H\hat{\beta}-h$, da unter der Nullhypothese erfüllter Restriktionen näherungsweise $d \sim N(0, \sigma^2 H(X'X)^{-1}H')$ gilt. Asymptotisch folgt also die Testgröße $T_W=d'(H(X'X)^{-1}H')^{-1}/\sigma^2$ einer Chi-Quadrat Verteilung mit g Freiheitsgraden.

Dieser recht einfach zu konstruierende Test ist in EViews verfügbar. Man erhält ihn aus dem Gleichungsmenü über View Coefficient_Tests Wald_Coefficient_Restrictions. Im geöffneten Panel muss man dann die Restriktionen in Gleichungsform eingeben, wobei $c(i)$ den jeweils i -ten Parameter bezeichnet. Als Resultat ergibt sich der folgende Output:



Wald Test:
Equation: EQ01

Test Statistic	Value	df	Probability
F-statistic	0.026985	(1, 140)	0.8698
Chi-square	0.026985	1	0.8695

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(2) - C(3) - C(4) - C(5)	0.077424	0.471323

Restrictions are linear in coefficients.

Der Wald-Test liefert also mit einem p-Wert von 0.87 keine Ablehnung der Homogenitätshypothese.

Die in EViews nicht direkt implementierten und hier nicht näher behandelten beiden anderen Tests sind übrigens asymptotisch äquivalent, die Teststatistiken sind also auch asymptotisch Chi-quadrat verteilt mit g Freiheitsgraden. Allerdings zeigt sich, dass bei linearen Modellen für kleine Stichproben gilt, dass $W \geq LR \geq LM$ gilt, der Wald-Test also am wenigsten an der Nullhypothese festhält.

Nun haben wir also gezeigt, dass die Homogenitätshypothese von unserer Schätzung nicht verletzt ist, trotzdem wäre es aber doch modellkonformer, könnten wir diese gleich bei der Schätzung mitberücksichtigen. Naheliegender wäre doch diese Bedingung durch Einsetzen eines der als Funktion der anderen Parameter ausgedrückten drei Preiskoeffizienten in die Gleichung. Nerlove verwendet dafür zum Beispiel den Brennstoffpreis p_f , also $a_f = r - a_l - a_k$. Nach Umformungen ergibt sich somit, was Nerlove Modell A nennt:

$$\log c - \log p_f = \log a_0 + \frac{1}{r} \log Y + \frac{a_l}{r} (\log p_l - \log p_f) + \frac{a_k}{r} (\log p_k - \log p_f) + \frac{1}{r} \log u.$$

oder

$$\log(c/p_f) = \log a_0 + \frac{1}{r} \log Y + \frac{a_l}{r} \log(p_l/p_f) + \frac{a_k}{r} \log(p_k/p_f) + \frac{1}{r} \log u.$$

Bei der Schätzung dieser Gleichung muss angenommen werden, dass für alle Versorger der Preis des Kapitals bekannt ist, und dass dieser Preis über die einzelnen Unternehmen schwankt. In Bezug auf die Finanzierungskosten erscheint diese Annahme unplausibel. Die Kreditzinssätze oder die Zinssätze von Unternehmensanleihen desselben Sektors sollten keine großen Unterschiede zeigen. Über unterschiedliche Abschreibungsraten auf das physische Kapital können jedoch durchaus unterschiedliche Kapitalkosten zwischen den Versorgern entstehen.

Die entsprechende OLS-Schätzung dieses restringierten Modells A liefert in EViews nun:

Dependent Variable: LOG(TC/PF)
 Method: Least Squares
 Sample: 1 145
 Included observations: 145

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.690789	0.884871	-5.301098	0.0000
LOG(Y)	0.720688	0.017436	41.33398	0.0000
LOG(PK/PF)	-0.007381	0.190736	-0.038698	0.9692
LOG(PL/PF)	0.592910	0.204572	2.898291	0.0044
R-squared	0.931585	Mean dependent var		-1.484195
Adjusted R-squared	0.930129	S.D. dependent var		1.482087
S.E. of regression	0.391762	Akaike info criterion		0.990874
Sum squared resid	21.64032	Schwarz criterion		1.072991
Log likelihood	-67.83836	F-statistic		639.9802
Durbin-Watson stat	1.015369	Prob(F-statistic)		0.000000

Dieses Resultat enthält nun implizit die Homogenitätsrestriktion. Der für uns relevante Schätzwert des Koeffizienten von log Y ist nahezu unverändert, was als Modellbestätigung gewertet werden kann. Zudem kann nun definitiv die Hypothese konstanter Skalenerträge abgelehnt werden.

Wir benennen diese Spezifikation mit eq02, weil wir später noch darauf zurückgreifen werden.

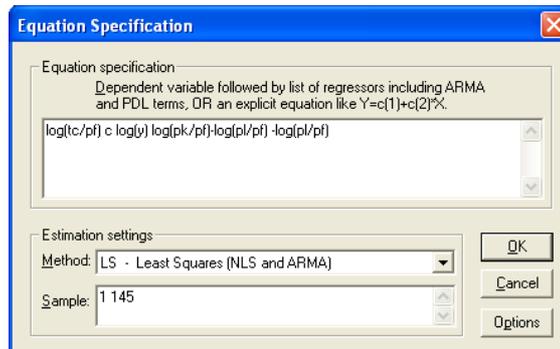
Der Outputtabelle können zwar nach wie vor die Schätzer und Standardfehler für a_i/r sowie für a_i/r direkt entnommen werden, der Schätzer für a_i/r muss jedoch aus der Nebenbedingung bestimmt werden. Durch Einführen der Restriktion haben wir die Matrix H und den Parametervektor β quasi in zwei Teile $H\beta = H_S\beta_S + H_N\beta_N = h$ zerlegt, wobei β_S nun die substituierten Parameter bezeichnet. Folglich ergibt sich für den Schätzer $\hat{\beta}_S = H_S^{-1}(h - H_N\hat{\beta}_N)$, was im Fall mit nur einer Bedingung einfach Einsetzen in die Restriktion, also hier $1+0.007-0.593 = 0.414$, ergibt.

Zur Bestimmung des Standardfehlers muss man auf die Kovarianzmatrix der Koeffizienten zugreifen, da ja $\text{Var}(\beta_S) = H_S^{-1}H_S \text{Var}(\beta_S) H_S' H_S^{-1}$. Diese kann man sich zuerst über VIEW Covariance_Matrix ansehen. Da in unserem Fall gilt, dass $h=1$, $H_N=1$ und $H_S=(0,0,1,1)$, bestimmt sich der Standardfehler für a_i/r einfach zu 0.099 aus:

Coefficient Covariance Matrix				
	C	LOG(Y)	LOG(PK/PF)	LOG(PL/PF)
C	0.782997	-0.003762	-0.159085	0.176333
LOG(Y)	-0.003762	0.000304	0.000213	-0.000533
LOG(PK/PF)	-0.159085	0.000213	0.036380	-0.034219
LOG(PL/PF)	0.176333	-0.000533	-0.034219	0.041850

```
rowvector(4) hsp
hsp.fill 0,0,1,1
matrix sterr_af=@sqrt(hsp*eq02.@cov*@transpose(hsp)v)
```

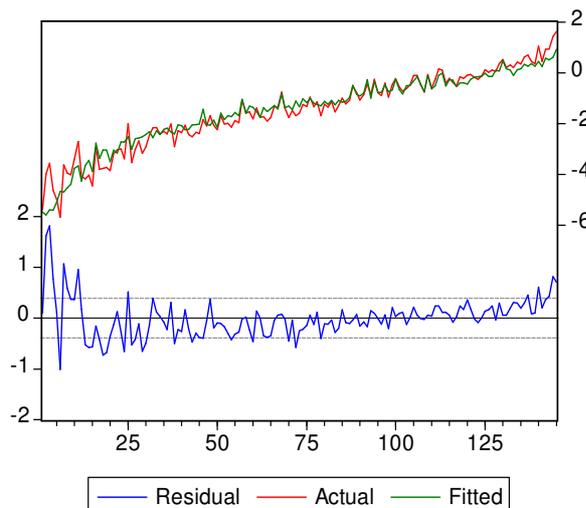
Hierzu musste in EViews zuerst der Zeilenvektor hsp definiert und dann entsprechend befüllt werden. Es ist auffällig, dass der skalare Wert sterr_af als 1x1 Matrix gespeichert werden muss. EViews kennt zwar den Datentyp Skalar, eine Überführung ist aber nicht möglich, was diesen Typus für praktische Zwecke ziemlich obsolet macht.



Man bemerke überdies, dass auch für die Bestimmung dieser Standardabweichung der Transformationstrick von Papke und Wooldridge (2005) angewendet werden kann, nämlich durch Durchführung der nebenstehenden Regression, da ja hier $\beta = (0,0,-1,-1)$.

4.3 Nichtlinearitäten, Feasible Generalized Least Squares

Damit haben wir nun praktisch ein modellkonformes Schätzergebnis. Es stellt

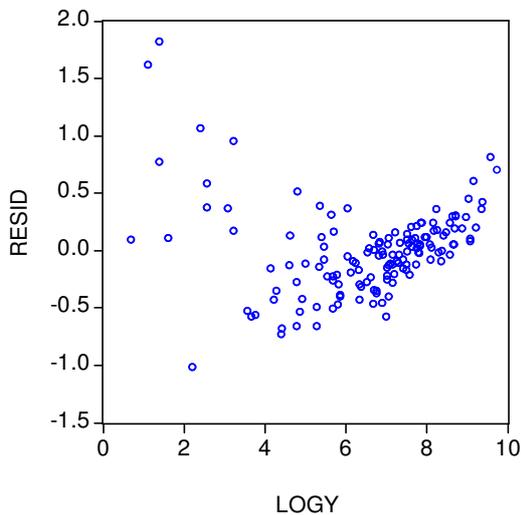


sich nur mehr die Frage wie zuverlässig dies ist. Eine erste Inspektion des Standardresidual-diagrammes in EViews, durch Klicken auf die Schaltfläche Resids im Fenster für equ02 zeigt verschiedentlich auffälliges Verhalten: die Reihe der Residuen weist eine leichte U-Form auf, die Streuung der Residuen scheint ebenfalls abzunehmen und die Reihe der beobachteten und der prognostizierten Werte zeigt einen Trend.

Letzteres kommt offenbar daher, dass die Beobachtungen dem Output nach geordnet sind und der Trend nur den Skaleneffekt widerspiegelt. Dieser Skaleneffekt hat natürlich auch Einfluss auf die Anpassungsgütemaße wie das R^2 . Es ist daher empfehlenswert, diese auf Basis der Durchschnittskosten TC/Y , die von dem Skaleneffekt befreit sind, zu evaluieren. Eine entsprechende Regression liefert gleich bleibende Koeffizienten, nur für $\text{LOG}(Y)$ nun natürlich $1/r-1$, das R^2 sinkt nun aber von 0.93 auf 0.65.

Dependent Variable: LOG(TC/Y/PF)
 Method: Least Squares
 Sample: 1 145
 Included observations: 145

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.690789	0.884871	-5.301098	0.0000
LOG(Y)	-0.279312	0.017436	-16.01956	0.0000
LOG(PK/PF)	-0.007381	0.190736	-0.038698	0.9692
LOG(PL/PF)	0.592910	0.204572	2.898291	0.0044
R-squared	0.653553	Mean dependent var	-8.040846	
Adjusted R-squared	0.646182	S.D. dependent var	0.658616	
S.E. of regression	0.391762	Akaike info criterion	0.990874	
Sum squared resid	21.64032	Schwarz criterion	1.072991	
Log likelihood	-67.83836	F-statistic	88.66281	
Durbin-Watson stat	1.015369	Prob(F-statistic)	0.000000	



Die erwähnte Nichtlinearität und Heteroskedastizität in den Residuen lässt sich dadurch aber nicht erklären, was sehr deutlich zu Tage tritt, wenn man die aus dieser Regression entstandenen Residuen gegen LOGY plottet.

Nein, hierzu muss man wieder etwas Modellierungsarbeit leisten. Die U-Form erklärt Nerlove dadurch, dass die Skalenerträge für verschiedene Größen von Unternehmen unterschiedlich sind. Er schlägt deshalb zur Behandlung des Problems zwei Lösungen vor: zuerst teilt er die Daten in 5 Größengruppen auf um separate Analysen durchzuführen (dies werden wir im nächsten Abschnitt besprechen). Dann erweitert er das Modell, in dem er als zusätzliche Variable das Quadrat von LOGY ins Modell aufnimmt, also

$$\log(c/p_f) = \log a_0 + \alpha \log Y + \frac{1}{2} \gamma \log^2 Y + \frac{a_l}{r} \log(p_l/p_f) + \frac{a_k}{r} \log(p_k/p_f) + \varepsilon$$

schätzt.

Dies motiviert sich aus der ad-hoc Annahme, dass die Skalenerträge invers proportional einer linearen Funktion der $\log Y$ entsprechen, also $r(Y) = (\alpha + \gamma \log Y)^{-1}$.

Die Nichtlinearität der Residuen soll dadurch abgefangen werden. Man bemerke jedoch, dass das Modell im statistischen Sinn ein Lineares bleibt – die Nichtlinearität beschränkt sich auf die Regressoren und betrifft nicht die Parameter – also weiterhin mittels OLS geschätzt werden kann. Wir speichern die folgende Gleichung unter eq03, weil wir auch sie später noch benötigen.

Dependent Variable: LOG(TC/PF)

Method: Least Squares

Sample: 1 145

Included observations: 145

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.764649	0.701727	-5.364837	0.0000
LOG(Y)	0.152547	0.061860	2.465992	0.0149
LOG(PK/PF)	0.074166	0.150016	0.494390	0.6218
LOG(PL/PF)	0.480586	0.161072	2.983663	0.0034
LOG(Y)^2/2	0.101028	0.010727	9.417823	0.0000
R-squared	0.958118	Mean dependent var	-1.484195	
Adjusted R-squared	0.956922	S.D. dependent var	1.482087	
S.E. of regression	0.307612	Akaike info criterion	0.513919	
Sum squared resid	13.24751	Schwarz criterion	0.616565	
Log likelihood	-32.25909	F-statistic	800.6873	
Durbin-Watson stat	1.665259	Prob(F-statistic)	0.000000	

Wie sieht das nun noch mit der Streuung aus? Die Fälle genau spezifizierter und völlig unspezifizierter Heteroskedastizität haben wir ja schon kennen gelernt. Hier kann man bestenfalls annehmen, dass die Varianz der Fehler mit steigendem Output fällt. Hayashi, 2000 schlägt die Spezifikation $\text{Var}(\varepsilon) \propto \lambda_0 + \lambda_1 Y^{-1}$ vor. Ist die Form der Varianz-Kovarianzmatrix V also nicht, wie im vorigen Kapitel, exakt bestimmbar, aber eine vorgegebene funktionale Struktur bekannt, könnte man sie – etwa aus den OLS-Residuen – vorschätzen, um anschließend mit $\hat{\beta}_{FGLS} = (X' \hat{V}^{-1} X)^{-1} X' \hat{V}^{-1} y$ den so genannten feasible GLS Schätzer (FGLS) zu bilden. Im Fall Heteroskedastizität geht es nur um die Diagonale von V und wir haben also ein einfaches zweistufiges Verfahren: zuerst regressiert man die Quadrate der Residuen auf die Kehrwerte des Outputs Y .

Dependent Variable: RESID03^2

Method: Least Squares

Sample: 1 145

Included observations: 145

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.056497	0.016285	3.469252	0.0007
1/Y	2.137700	0.260265	8.213564	0.0000
R-squared	0.320544	Mean dependent var	0.091362	
Adjusted R-squared	0.315793	S.D. dependent var	0.228877	
S.E. of regression	0.189320	Akaike info criterion	-0.477059	
Sum squared resid	5.125414	Schwarz criterion	-0.436001	
Log likelihood	36.58678	F-statistic	67.46264	
Durbin-Watson stat	2.036631	Prob(F-statistic)	0.000000	

Nun kann man sich über $v_0 = \sqrt{0.056497 + 2.137700/Y}$ die geschätzten Standardabweichungen bestimmen und dann neu mittels WLS mit GewichtungsvARIABLE $1/v_0$ schätzen, was aber hier sowohl bei den Parameterschätzern als auch in Bezug auf die Varianzstabilisation nur sehr geringe Effekte hat.

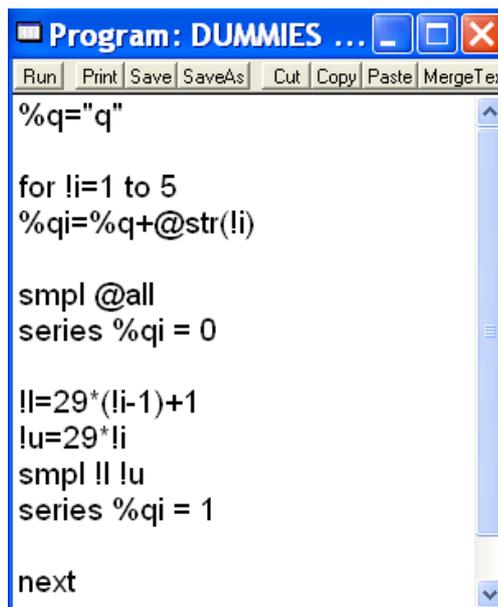
4.4 Chow-Test, Multikollinearität, Ridge Regression

Kommen wir nun zu der von Nerlove zusätzlich durchgeführten, bereits erwähnten Aufteilung der Daten in Größengruppen. Diese lässt sich einfach bewerkstelligen, da der Datensatz bereits der Outputgröße y nach geordnet ist, siehe auch die Variable order. Es würde also genügen, jeweils den sample Bereich um 29 Beobachtungen zu verschieben (so erhalten wir aus den 145 Beobachtungen fünf Gruppen) und separate Analysen durchzuführen.

Das klingt aber doch sehr umständlich. Wir können doch auch durch geeignetes Setzen von Dummyvariablen alles in einer Gleichung schätzen.

Das ist richtig, aber auch etwas mühsam, da wir fünf Dummyvariablen erzeugen müssen, die in den entsprechenden Gruppen jeweils den Wert 1, sonst 0 aufweisen.

Per „Hand“ wäre das zu viel Arbeit. Es gibt jedoch die Möglichkeit in Eviews kleine Programme zu definieren, die solche Aufgaben erledigen. Über New_Program öffnet sich ein Programmfenster. Hier können Befehlssequenzen wie in der Befehlszeile eingegeben werden, welche dann bei der Programmausführung (durch Klicken auf Run) hintereinander exekutiert werden. So kann auch Komplexeres erledigt werden, nebenstehendes Fenster enthält z.B. den Code zur Erzeugung der gewünschten fünf Dummyvariablen mit den Namen q1 bis q5.

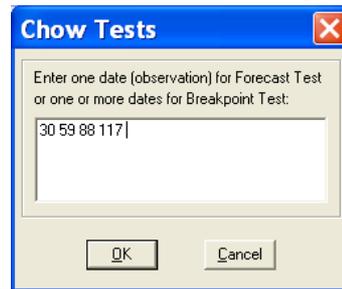


```
Program: DUMMIES ...
Run Print Save SaveAs Cut Copy Paste MergeText
%q="q"
for !i=1 to 5
%qi=%q+@str(!i)
smpl @all
series %qi = 0
!!=29*(!i-1)+1
!u=29*!i
smpl !! !u
series %qi = 1
next
```

Sind diese Dummyvariablen erzeugt, braucht man Sie nur mit den gewünschten Regressoren inklusive Interzept multiplizieren und diese dann in die Gleichung aufnehmen. Der entstandene Output enthält dann die entsprechenden Schätzer für alle Größengruppen. Durch Manipulation mit den Dummyvariablen können jetzt auch Hypothesen zur Rechtfertigung der getroffenen Aufteilung, etwa mittels der üblichen Wald- und F-Tests durchgeführt werden.

Ein sehr einfache Version lässt sich ohne Generierung der Dummies bereits im Vorhinein in Eviews bestimmen: der so genannte Chow-Bruchpunkt-Test überprüft einfach die Hypothese, ob sich die Koeffizienten des Modells in den einzelnen Untergruppen signifikant voneinander unterscheiden. Dieser spezifische Wald-Test lässt sich in Eviews im Gleichungsfenster über View-Stability_Tests-Chow-Breakpoint-Test aufrufen. Im Dialogfenster müssen anschließend

die Startwerte für die Unterteilungen angegeben werden (hier siehe die nebenstehenden Einträge). Das Resultat gibt dann an, ob in den in den fünf Teilgruppen zu je 29 Beobachtungen die gleiche lineare Struktur vorliegt. Aufgrund der kleinen p-Werte für Gleichung eq02 kann ohne Zweifel diese Nullhypothese abgelehnt und bestätigt werden, dass bei den 145 Unternehmen unterschiedliche Skaleneffekte vorliegen (für Gleichung eq03 ergibt sich jedoch eine Ablehnung des Chow-Tests).



Chow Breakpoint Test: 30 59 88 117

F-statistic	5.974731	Probability	0.000000
Log likelihood ratio	82.36260	Probability	0.000000

Es bietet sich also nun zumindest eine Analyse des Modells eq02 für die 5 Untergruppen mittels Dummy-Regression an:

Dependent Variable: LOG(TC/PF)

Method: Least Squares

Sample: 1 145

Included observations: 145

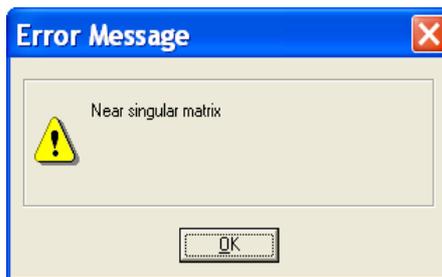
Variable	Coefficient	Std. Error	t-Statistic	Prob.
Q1	-3.343348	1.652646	-2.023027	0.0452
Q1*LOG(Y)	0.400290	0.044370	9.021601	0.0000
Q1*LOG(PK/PF)	-0.081356	0.371103	-0.219227	0.8268
Q1*LOG(PL/PF)	0.615173	0.383176	1.605461	0.1109
Q2	-6.488974	1.825601	-3.554432	0.0005
Q2*LOG(Y)	0.658151	0.150263	4.380006	0.0000
Q2*LOG(PK/PF)	0.377936	0.357290	1.057785	0.2922
Q2*LOG(PL/PF)	0.093800	0.354396	0.264675	0.7917
Q3	-7.332942	2.671684	-2.744689	0.0069
Q3*LOG(Y)	0.938279	0.313146	2.996304	0.0033
Q3*LOG(PK/PF)	0.250008	0.295837	0.845087	0.3997
Q3*LOG(PL/PF)	0.402258	0.315467	1.275119	0.2046
Q4	-6.546049	3.025269	-2.163791	0.0324
Q4*LOG(Y)	0.912044	0.279177	3.266899	0.0014
Q4*LOG(PK/PF)	0.093352	0.426200	0.219033	0.8270
Q4*LOG(PL/PF)	0.506958	0.486941	1.041106	0.2998
Q5	-6.714258	2.181091	-3.078394	0.0026
Q5*LOG(Y)	1.044390	0.135455	7.710217	0.0000
Q5*LOG(PK/PF)	-0.289436	0.364519	-0.794021	0.4287
Q5*LOG(PL/PF)	0.602588	0.411281	1.465149	0.1454
R-squared	0.961233	Mean dependent var		-1.484195
Adjusted R-squared	0.955340	S.D. dependent var		1.482087
S.E. of regression	0.313208	Akaike info criterion		0.643546
Sum squared resid	12.26243	Schwarz criterion		1.054130
Log likelihood	-26.65706	Durbin-Watson stat		1.774708

Dabei zeigt sich, dass bis auf die höchste Gruppe ($\hat{r}=1/1.044=0.96$) alle anderen steigende Skalenerträge aufweisen.

Man bemerke, dass auch die Spezifikation eq03 mit dem zusätzlichen Regressor $\log Y^2/2$ in der Dummyversion geschätzt werden kann. Die aus dieser Gleichung ableitbaren Skalenerträge für die verschiedenen Firmen an den Medianen der Größengruppen entsprechen aber nicht den in Nerloves Tabelle 8 wiedergegebenen Werten, weil dieser fälschlicherweise $r(Y) = (\alpha + 1/2)\log Y$ zur Berechnung verwendete.

Bei einer Verwendung von Dummies wie oben ist es auch immer nötig, darauf Acht zu geben, nicht in die sogenannte Dummyvariablenfalle zu tappen, und damit die Kleinst-Quadrateschätzung unmöglich zu machen. Dies geschieht dann, wenn man die Dummies so kodiert, beziehungsweise ins Modell aufnimmt, dass sie untereinander oder mit anderen Regressoren Linearkombinationen bilden.

In obigem Beispiel wäre dies etwa bei der zusätzlichen Aufnahme eines Interzepts der Fall, da ja dann einfach $C=Q1+Q2+Q3+Q4+Q5$. EViews liefert allerdings in einem solchen Fall die adäquate nebenstehende Fehlermeldung.



Leider kann dieses Phänomen – der sogenannten perfekten Kollinearität – aber nicht nur bei der Verwendung von Dummies sondern ganz generell auftreten, und zwar immer dann, wenn die Designmatrix $X'X$ nicht vollen Rang aufweist, das heißt singular ist. In der Regel passiert das jedoch nur, wenn bei der Festlegung der Regressoren ein Designfehler begangen wird. Allerdings tritt eine abgeschwächte Form dieses Problems durchaus in der Praxis auf, nämlich dann, wenn einzelne Regressoren annähernd kollinear sind, also stark miteinander (multivariat) korrelieren. Dieser Zustand erhielt in der Ökonometrie den Namen Multikollinearität.

Eigentlich braucht man sich vor diesem Phänomen wohl nicht zu fürchten, da die Optimalitätseigenschaften der Kleinst-Quadrate Schätzung, also das Faktum dass diese beste lineare unverzerrte Schätzer liefert, unberührt bleibt. Allerdings kann Multikollinearität dazu führen, dass dieser Schätzer nur mehr sehr ungenau ist, das heißt, seine Varianz sehr groß wird.

Am besten lässt sich dieser Effekt an Hand eines zwei-Variablen Modells ohne Interzept illustrieren. Hier gilt nämlich, dass man den Schätzer für die Varianz des Kleinst-Quadrateschätzers $\hat{\beta}_i$ für $i=1,2$ schreiben kann als $\text{Var}(\hat{\beta}_i) = \{(1-R^2)m_y\}/\{n(1-r^2)m_i\}$, wobei r den Korrelationskoeffizienten der Regressoren und m_y und m_i die zweiten zentralen Stichprobenmomente von Regressand und Regressoren bezeichnet. Diese Darstellung zeigt klar das ceteris paribus potenziell explosive Ansteigen dieser Varianz bei wachsendem r .

Aber nicht nur die geschätzte, auch die tatsächliche Varianz des Schätzers ist selbstverständlich betroffen, denn man kann sie allgemein schreiben als

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma_\varepsilon^2}{(1-R_j^2) \sum_{i=1}^T (x_{ij} - \bar{x}_j)^2},$$

wobei R_j^2 wieder das Bestimmtheitsmaß aus einer Hilfsregression der einzelnen Regressoren x_j auf alle übrigen Regressoren bezeichnet. Für diese Varianzen sind also mehrere Faktoren, wie die Störgrößenvarianz, die Varianzen der Regressoren, die Zahl der Beobachtungen (über die Summierung im Nenner) verantwortlich; die auf die Multikollinearität zurückzuführenden Faktoren sind aber

$$VIF_j = (1 - R_j^2)^{-1}.$$

Diese werden Varianzinflationsfaktoren genannt und dienen, wie auch die R_j^2 's selbst, als Indikator für das Ausmaß der den jeweiligen Regressor betreffenden Multikollinearität.

Es ist sicher hilfreich solche Indikatoren zu haben, aber wie entscheiden wir, ob wir es diesbezüglich mit einem gravierenden Problem zu tun haben?

Nun, in der ökonometrischen Praxis geistert die durchaus umstrittene Faustregel herum, die VIF's sollten kleiner als 10 sein, also die Varianz um weniger als das Zehnfache aufgebläht werden. Für die R_j^2 's hat Lawrence Klein übrigens die Faustregel aufgestellt, sie mögen den Wert des Gesamtbestimmtheitsmaßes R_j^2 nicht überschreiten.

Die Varianzinflationsfaktoren werden bei einigen Programmen im Standardregressionsoutput ausgegeben, in EViews müssen sie leider auf recht umständliche Art erzeugt werden. Wir verwenden in der Folge der Einfachheit halber die Nerlove Gleichung eq03 ohne Dummies und Heteroskedastiekorrektur. Zuerst generiert man sich eine Regressorengruppe (mit Namen groupx). Dafür gibt es an und für sich im Eviews Gleichungoutput über Procs-Make_Regressor_Group eine direkte Möglichkeit, doch enthält die entstandene Gruppe auch noch den hier nicht benötigten Regressanden, den man aber über View-Group_Members (oder Spec) wieder aus der Gruppe entfernen kann. Weil wir in der ersten Spalte wegen des Interzepts später aber ohnehin ein Element zu viel haben, können wir ihn vorerst belassen. Danach erstellt man die Kovarianzmatrix der Regressoren über @cov(groupx). Die Varianzen aus der Diagonale dieser Matrix müssen nun mit den quadrierten Standardfehlern der Regression multipliziert werden. Weil in Eviews punktweise Multiplikationen bei Vektoren und Matrizen nicht möglich sind, müssen wir recht umständlich mit äußeren Produkten und Diagonalextraktionen arbeiten:

```
vector v1=@getmaindiagonal(eq03.@stderrs*@transpose(eq03.@stderrs))
vector v2=@getmaindiagonal(@cov(groupx))
vector vif=@getmaindiagonal(v1*@transpose(v2))/eq03.@se^2*@regobs
```

Man beachte, dass zuletzt noch die Normierung SSE^2/T durchgeführt werden muss.

Aus dem nebenstehenden Resultat (unter Ignorieren der ersten, bedeutungslosen Zeile) erkennen wir, dass sowohl für die Variable logY als auch $\log^2 Y/2$ die Faustregel Multikollinearität anzeigt (Klein's Regel wird übrigens für diese beiden Variablen knapp nicht verletzt).

VIF	
	C1
	Last updated: ...
R1	1646.037
R2	21.30657
R3	4.433506
R4	4.525975
R5	21.39046

Es ist zu bemerken, dass das Verhältnis der geschätzten Varianzen (einmal im Modell mit nur der jeweiligen Variablen plus Interzept, dann im vollen Modell) diesen Werten selbstverständlich nicht exakt entspricht. Für den Term $\log^2 Y/2$ etwa erhalten wir einen Faktor der quadrierten Standardfehler von $(0.010727^2/0.002722^2)=15.53$.

Als weiteres gebräuchliches Indiz für Multikollinearität hat sich übrigens der so genannte Konditionsindex (CI) durchgesetzt. Dieser errechnet sich als die Quadratwurzel aus dem Verhältnis des größten zum kleinsten Eigenwert der Designmatrix. Er lässt sich am einfachsten über die Kovarianzmatrix der Koeffizienten bestimmen, nämlich aus

vector ev=@eigenvalues(@inverse(eq03.@coefcov))

welches einen Wert von $CI=\sqrt{(1102187/1.867602)} \approx 768$ liefert, weit über dem Faustregelschwellwert von hier 30.

Wir haben nun festgestellt, dass in unserer Gleichung vermutlich das Problem der Multikollinearität vorliegt. Was tun wir also dagegen?

Hier kann ich mit einer Hierarchie von Vorschlägen aufwarten:

1. Nichts tun! Weil es sich bei Multikollinearität in erster Linie nicht um ein Modellierungs- oder Schätzproblem handelt: es kann zum Beispiel durchaus sein, dass bestimmte Linearkombinationen der Parameter außerordentlich gut geschätzt werden können. Für den Zweck der Vorhersage ist ihr Vorhandensein beziehungsweise Nichtvorhandensein völlig unbedeutend. Eng verwandt mit diesem Ausweg ist natürlich:
2. Mehr Daten! Es handelt sich hier um ein Problem der Datenqualität, welches formal völlig dem der zu kleinen Stichproben entspricht. Arthur Goldberger hat um darauf hinzuweisen, deshalb auch den etwas euphemistischen Begriff „Micronumerosity“ eingeführt. Oft ist dies Erheben zusätzlicher Daten aber nicht durchführbar, deshalb verwendet man manchmal
3. Schätzer aus anderen Studien (Metaanalysen), bzw. formalisierte Vorinformation (Bayesverfahren). Wenn selbst dies nicht zur Verfügung steht, kann man auf verschiedene Arten die Komplexität des Modells reduzieren, nämlich durch
4. Formalisieren der Beziehungen zwischen den Regressoren (Erstellung und Schätzung von Gleichungssystemen) oder
5. Spezifizieren der Beziehungen zwischen den Parametern (Einführung von Restriktionen und bestimmten Lagformen), schließlich
6. das Bilden von Transformationen (zum Beispiel Hauptkomponenten) der Regressoren.

Der Versuchung Variable wegzulassen sollte man tunlichst widerstehen, sind diese nämlich von Relevanz, kann dies zu erheblichen Verzerrungen der Parameterschätzer führen.

Zuletzt sei noch eine „neutrale“ Maßnahme erwähnt, die zuletzt verstärkt angewandt wird, so genannte „shrinkage-estimators“. Die einfachste Variante davon ist die von Hoerl und Kennard (1970) eingeführte ridge regression. Dabei wird einfach bei der Bestimmung der Schätzer zur Regularisierung der

Designmatrix $X'X$ ein Vielfaches der Einheitsmatrix dazu addiert, welches automatisch die Kollinearität der Regressoren reduziert.

Die Schätzwerte ergeben sich dann also aus:

$$\hat{\beta}_R = (X'X + kI)^{-1} X'Y.$$

Wie bestimmt man aber den konkreten Wert des Vielfachen k ?

Dies ist genau das Problem bei der Sache. Man kann natürlich das k gerade so klein wählen, so dass die Multikollinearitätsfaustregeln gerade nicht mehr verletzt werden. Dann hat man zwar dieses Problem in den Griff gekriegt, aber man weiß natürlich nicht, wie viel arbiträrer Einfluss durch die Einheitsmatrix in die Schätzung letztendlich eingeht.

Gut, wir können aber doch das k dann innerhalb dieses Bereichs variieren, um eine Vorstellung dafür zu gewinnen, ob die eingesetzte Korrektur zu zu starken Verzerrungen führt, beziehungsweise ab wann die Parameterschätzungen auffälligen Veränderungen unterworfen sind.

Dazu brauchen wir aber eine Möglichkeit diese Schätzer in EViews zu bestimmen. Es zeigt sich, dass die Verwendung der Ridge-regression gleichbedeutend mit der Aufnahme der nichtlinearen Parameterrestriktion $\beta'\beta = c(k)$ ist. Durch Auflösen nach dem Interzept und anschließendem Einsetzen ergibt sich das neue, nun nichtlineare Modell zu

$$Y = \text{sign}(\beta_0) \sqrt{c(k) - \beta_{\{0\}}' \beta_{\{0\}}} + X_{\{0\}} \beta_{\{0\}},$$

wobei der Index $\{0\}$ für das Weglassen des Interzepts steht. Dies lässt sich in EViews nun direkt mit der (nichtlinearen) Kleinst-Quadrate Methode schätzen, wir benötigen hier nur die algebraische Modellschreibweise, bei der $c(1)$, $c(2)$, ... für die Parameter stehen. Setzen wir zum Beispiel $c(k)=1.5$ für unser Modell aus Gleichung eq03 ergibt sich dann:

Dependent Variable: LOG(TC/PF)

Method: Least Squares

Sample: 1 145

Included observations: 145

Convergence achieved after 94 iterations

LOG(TC/PF) = -@SQRT(1.5-C(1)^2-C(2)^2-C(3)^2-C(4)^2)+C(1)* LOG(Y)
 +C(2)*LOG(PK/PF)+C(3)*LOG(PL/PF)+C(4)*LOG(Y)^2/2

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.126783	0.048722	-2.602153	0.0103
C(2)	-0.484111	0.137611	-3.517969	0.0006
C(3)	1.032177	0.153576	6.720935	0.0000
C(4)	0.143087	0.009481	15.09227	0.0000
R-squared	0.946138	Mean dependent var		-1.484195
Adjusted R-squared	0.944992	S.D. dependent var		1.482087
S.E. of regression	0.347604	Akaike info criterion		0.751693
Sum squared resid	17.03682	Schwarz criterion		0.833810
Log likelihood	-50.49773	Durbin-Watson stat		1.501916

Der Wert $c(k)$ ist umgekehrt proportional zu k , das bedeutet, dass jede Reduktion von $c(k)$ zu einer Reduktion der Multikollinearität führt, aber auch zu einer potenziell stärkeren Veränderung der Schätzwerte, wie in diesem Fall.

Deshalb wird die Ridge-Regression oft auch als zu rigide kritisiert, siehe etwa Jensen und Ramirez (2008). Insbesondere erlaubt diese Technik nicht, dass störende Variable eliminiert werden, was zum Beispiel eine Variante, das so genannte „lasso“ von Tibshirani (1996) leistet. Hier lautet die nötige Parameterrestriktion einfach $\sum |\beta_i| = c(k)$.

Nach entsprechender Transformation und der Wahl $c(k)=3.6$ ergibt sich in unserem Fall dann:

Dependent Variable: LOG(TC/PF)

Method: Least Squares

Date: 03/20/08 Time: 15:18

Sample: 1 145

Included observations: 145

Convergence achieved after 4 iterations

LOG(TC/PF) = -(3.6*@ABS(C(1)))-@ABS(C(2))-@ABS(C(3))-@ABS(C(4))
 +C(1)* LOG(Y)+C(2)*LOG(PK/PF)+C(3)*LOG(PL/PF)+C(4)*LOG(Y)^2/2

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.100061	0.055717	1.795878	0.0747
C(2)	-0.154582	0.088768	-1.741419	0.0838
C(3)	0.739026	0.085006	8.693801	0.0000
C(4)	0.108797	0.009991	10.88994	0.0000
R-squared	0.957058	Mean dependent var		-1.484195
Adjusted R-squared	0.956144	S.D. dependent var		1.482087
S.E. of regression	0.310375	Akaike info criterion		0.525124
Sum squared resid	13.58286	Schwarz criterion		0.607241
Log likelihood	-34.07153	Durbin-Watson stat		1.682427

Dieses Ergebnis entspricht inhaltlich viel stärker der OLS-geschätzten Gleichung eq03. Zudem wurden die Multikollinearitätsmaße stark reduziert; der größte VIF-wert beträgt etwa nur noch 13.08.

Übungsaufgaben

- Erstellen Sie eine Schätzung, bei der nur die Konstante und der Outputkoeffizient über die fünf Gruppen variieren (Nerlove nennt dies neutrale Variationen in Skalenerträgen). Testen Sie die zugrunde liegende Restriktion.
- Da die Datengrundlage für den Preis des Kapitals in der Stichprobe von Nerlove nicht besonders gut ist, schätzt Nerlove auch eine zweite Variante der Kostenfunktion. Unter der Annahme, dass der Preis des Kapitals für alle Versorger gleich hoch ist, kann die Restriktion auf die Koeffizienten des Outputs und der Preise dazu genutzt werden, die Elastizität des Outputs auf den Kapitalinput zu schätzen. Unter der Annahme identischer Kapitalpreise sieht die Kostenfunktion folgendermaßen aus:

$$\log c = k + \frac{1}{r} \log Y + \frac{a_l}{r} \log p_l + \frac{a_f}{r} \log p_f + \frac{1}{r} \log u$$

wobei die Konstante nun $k = \log a_0 + (a_k/r)p_k$ ist. Diese Gleichung nennt Nerlove Modell B.

- Christensen und Greene (1976) analysierten den US-Elektrizitätsmarkt erneut. Unter Benützung des Workfiles 04ch&h.wf1 versuchen Sie deren Analyse nachzuvollziehen. Erstellen Sie insbesondere entsprechende Durchschnittskostenkurven.

Literaturhinweise:

Christensen, L.R. und Greene, W.H., “Economies of Scale in U.S. Electric Power Generation”, *The Journal of Political Economy* 84(4), 655-676, 1976.

Cobb, C. und Douglas, P.H., „A Theory of Production“, *American Economic Review, Supplement* 18, 139-165, 1928.

Dias Curto, J. und Castro Pinto, J. “New Multicollinearity Indicators in Linear Regression Models”, *International Statistical Review* 75(1), 114-121, 2007.

Hackl, P. „Einführung in die Ökonometrie“, Pearson Studium, 2004.

Hayashi, F., “Econometrics”, Princeton University Press, 2000.

Hoerl, A.E. und Kennard, R.W., “Ridge Regression: Biased Estimation for Nonorthogonal Problems”, *Technometrics* 12, 55-67, 1970.

Jensen, D.R. und Ramirez, D.E., “Anomalies in the Foundations of Ridge Regression”, *International Statistical Review* 76(1), 89-105, 2008.

Moeller, S.B., Schlingemann, F.P., Stulz, R.M., "Wealth Destruction on a Massive Scale? A Study of Acquiring-Firm Returns in the Recent –Merger Wave", National Bureau of Economic Research, Working Paper No. 10200, Cambridge MA, 2004.

Nerlove, M., Analysis of Costs, in Christ, C.F. et al., “Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld”, Stanford University Press, Stanford, 1963.

Papke, L.E., Wooldridge, J.M., “A Computational Trick for Delta-Method Standard Errors”, *Economics Letters* 86, 413-417, 2005.

Tibshirani, R., “Regression Shrinkage and Selection via the Lasso”, *Journal of the Royal Statistical Society Series B*, 58(1) 267-288, 1996.

Tichy, G., "What do we Know About Success and Failure of Mergers?", *Journal of Industry, Competition and Trade* 1, 347-394, 2001.

Varian, H.R., Intermediate Microeconomics – A Modern Approach, 3rd Ed., W.W. Norton and Company, New York, 1993.

Varian, H.R., “Microeconomic Analysis”, W.W. Norton & Company, New York, 1992.